

Application Serial No.: 10/000,271
Amendment and Response to January 10, 2005 Non-Final Office Action
and Telephone Interview of August 18, 2005

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. (Cancelled)

2. (Currently amended) A method according to Claim 45, wherein the step of determining the third value comprises:

determining, for each of the first plurality of data fields and respective ones of the second plurality of data fields, a fourth value based on a mean of a first value determined for one of the first plurality of data fields and a second value determined for a respective one of the second plurality of data fields; and

summing the determined fourth values.

3. (Currently amended) A method for determining a value representing a difference between a first record comprising a first plurality of data fields and a second record comprising a second plurality of data fields, each of the first plurality of data fields corresponding to a respective one of the second plurality of data fields, the method comprising:

for each of the first plurality of data fields, determining a first value representing a difference between data specified in the data field and data specified in a respective one of the second plurality of data fields;

for each of the second plurality of data fields, determining a second value representing a difference between data specified in the data field and data specified in a respective one of the first plurality of data fields;

determining a third value representing a difference between the first record and the second record based on the determined first and second values; and

Application Serial No.: 10/000,271
Amendment and Response to January 10, 2005 Non-Final Office Action
and Telephone Interview of August 18, 2005

identifying whether the first and second records are duplicates based on the determined third value, wherein the determining and identifying is provided by a processor.

A method according to Claim 1, wherein the step of determining of the third value comprises:

determining a sum of the determined first values and the determined second values; and

dividing the sum by two.

4. (Currently amended) A method according to Claim 15, wherein the step of determining of the first value and the step of determining of the second value comprise identical operations steps performed with respect to different inputs.

5. (Currently amended) A method for determining a value representing a difference between a first record comprising a first plurality of data fields and a second record comprising a second plurality of data fields, each of the first plurality of data fields corresponding to a respective one of the second plurality of data fields, the method comprising:

for each of the first plurality of data fields, determining a first value representing a difference between data specified in the data field and data specified in a respective one of the second plurality of data fields;

for each of the second plurality of data fields, determining a second value representing a difference between data specified in the data field and data specified in a respective one of the first plurality of data fields;

determining a third value representing a difference between the first record and the second record based on the determined first and second values; and

identifying whether the first and second records are duplicates based on the determined third value, wherein the determining and identifying is provided by a processor.

Application Serial No.: 10/000,271
Amendment and Response to January 10, 2005 Non-Final Office Action
and Telephone Interview of August 18, 2005

A method according to Claim 1, wherein the step of determining of the first value comprises:

determining an asymmetric spelling distance as a normalized cost for converting first input data to second input data via a sequence of operations; and

wherein the step of determining the second value comprises:

determining an asymmetric spelling distance as a normalized cost for converting second input data to first input data via the sequence of operations

6. (Currently amended) A method according to Claim 5, wherein, in the step of determining of the first value, the first input data is data specified in one of the first plurality of data fields and the second input data is data specified in a respective one of the second plurality of data fields, and

wherein, in the step of determining of the second value, the first input data is data specified in one of the second plurality of data fields and the second input data is data specified in a respective one of the first plurality of data fields.

7. (Currently amended) A method according to Claim 15, further comprising:

converting numerical data specified in the one or more of the first plurality of data fields and the second plurality of data fields to text data.

8. (Currently amended) A method according to Claim 15, wherein the first plurality of data fields and the second plurality of data fields include only those fields of the first record and the second record that specify data that is not identical to data specified in a respective field.

9 - 32. (Canceled)

33. (Currently amended) An apparatus-medium according to Claim 342, wherein the instruction-step to determine the first value and the instructions step-to

Application Serial No.: 10/000,271
Amendment and Response to January 10, 2005 Non-Final Office Action
and Telephone Interview of August 18, 2005

determine the second value comprise identical steps performed with respect to different inputs.

34. (Currently amended) An apparatus medium according to Claim 32
storing processor-executable instructions thereon to determine a value representing a difference between a first record comprising a first plurality of data fields and a second record comprising a second plurality of data fields, each of the first plurality of data fields corresponding to a respective one of the second plurality of data fields, the instructions comprising:

instructions to determine, for each of the first plurality of data fields, a first value representing a difference between data specified in the data field and data specified in a respective one of the second plurality of data fields;

instructions to determine, for each of the second plurality of data fields, a second value representing a difference between data specified in the data field and data specified in a respective one of the first plurality of data fields;

instructions to determine a third value representing a difference between the first record and the second record based on the determined first and second values; and

instructions to identify whether the first and second records are duplicates based on the determined third value,

wherein the instructions step-to determine the first value comprises:

instructions a-step-to determine an asymmetric spelling distance as a normalized cost for converting first input data to second input data via a sequence of operations;
and

wherein the instructions step-to determine the second value comprises:

instructions a-step-to determine an asymmetric spelling distance as a normalized cost for converting second input data to first input data via the sequence of operations.

35. (Currently amended) An apparatus medium according to Claim 34,
wherein, in the instructions step-to determine the first value, the first input data is data

Application Serial No.: 10/000,271
Amendment and Response to January 10, 2005 Non-Final Office Action
and Telephone Interview of August 18, 2005

specified in one of the first plurality of data fields and the second input data is data specified in a respective one of the second plurality of data fields, and

wherein, in the instructions step to determine the second value, the first input data is data specified in one of the second plurality of data fields and the second input data is data specified in a respective one of the first plurality of data fields.

36. (Currently amended) An apparatus ~~medium~~ according to Claim 342, wherein the first plurality of data fields and the second plurality of data fields include only those fields of the first record and the second record that specify data that is not identical to data specified in a respective field.

37 - 38. (Canceled)

39. (Currently amended) A data warehouse according to Claim 38, wherein the instructions adapted to be executed by the processor to determine the first value and to determine the second value comprise identical steps performed with respect to different inputs.

40. (Currently amended) A data warehouse according to Claim 38,
comprising:

a processor; and

a storage device in communication with the processor and storing instructions
adapted to be executed by the processor to:

determine, for each of a first plurality of data fields of a first record, a first
value representing a difference between data specified in the data field and data
specified in a respective one of a second plurality of data fields of a second
record.

determine, for each of the second plurality of data fields, a second value
representing a difference between data specified in the data field and data
specified in a respective one of the first plurality of data fields,

Application Serial No.: 10/000,271
Amendment and Response to January 10, 2005 Non-Final Office Action
and Telephone Interview of August 18, 2005

determine a third value representing a difference between the first record
and the second record based on the determined first and second values; and

identify whether the first and second records are duplicates based on the
determined third value., wherein the instructions adapted to be executed by the
processor to determine the first value comprise instructions adapted to be executed by
the processor to:

determine an asymmetric spelling distance as a normalized cost for converting
first input data to second input data via a sequence of operations; and

wherein the instructions adapted to be executed by the processor to determine
the second value comprise instructions adapted to be executed by the processor to:
determine an asymmetric spelling distance as a normalized cost for converting the
second input data to the first input data via the sequence of operations.